

## **Data and Software Management Plan**

The PI, co-PI, and the students associated with our team are all strongly committed to compliance with US Census Bureau policies on the preservation, dissemination, and sharing of research data. However, it must be recognized that there could be some overlap between the proposed project and other projects of the PIs. These overlapping parts will be governed directly by the investigators' institutions and/or external funding organizations. The team will pursue activities designed to enable and encourage our investigators and users to comply with US Census Bureau data management policies. The PI will have primary responsibility for developing a set of policies on data management and overseeing their implementation.

### **Types of Data**

We expect to use public data extensively. In addition, we also plan to use data accessed from the Connecticut Health Information Exchange, UConn Police Department, as well as the Connecticut Transportation Safety Research Center. We would greatly appreciate data from the US Census Bureau.

### **Data and Metadata Standards**

Data formats will follow standard protocols based on the type of data. For example, records can be stored as text files. Relevant publications describing the data and analysis will be stored as .pdf files.

### **Policies for access and sharing and provisions for appropriate protection/privacy**

Relevant data will be made publicly available in a timely manner.

### **Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements**

We do not anticipate any intellectual property, or ethical or privacy issues stemming from this research.

### **Policies and provisions for re-use, re-distribution**

Access to most of the data produced by this research will generally be unrestricted, with the caveat described above. The data could be reused in comparative studies. We do not see any ethical or intellectual property reasons preventing sharing or reuse of the data.

### **Plans for archiving and Preservation of access**

Data will be submitted with pertinent metadata to archives such as GitHub. We will also store relevant data in the web site to be developed for this project.

### **General Data Management Policies and Activities**

We intend to take several steps to educate our users and encourage compliance with US Census Bureau's and our data management policies. First, each participant of the project (including investigators, students, and users) will be required to agree to a user responsibility document (URD). Among other things, our URD will notify users that they are obligated to: (a) protect shared resources and all data; (b) abide by US Census Bureau policies and other laws or regulations that pertain to intellectual property created; and (c) properly acquire and obtain proper licenses for all software and other data used. In addition our URD will specify that all activities undertaken in

the project are intended for publication in the open literature and that software developed is subject to copyright restrictions according the US Census Bureau General Grant Conditions or other applicable documents.

Second, the PI and co-PIs will work with participating students and users at their own institutions to educate and assist them in using both our data preservation and management facilities and those of their own institutions. Finally, the PIs will develop a data management guide that we will publish on our website and distribute to all users. In this regard, we note that our institutions take data management very seriously.

### **Data and Software Management Policies and Activities**

Given the theme of the project, we anticipate that our investigators will handle very diverse data and software. We have discussed data management above. We will assist users in preserving their software in much the same way as we do for their data. In addition, the participants will share software and tools they create with the broader research community. Any software and tools developed by the PIs will be viewed as open source and will be made available to others upon request, subject to US Census Bureau copyright restrictions that may apply. In addition, the PI and co-PIs are specifically committed to wide dissemination of the software and tools developed, through a combination of publication and open-source downloads from a website that we will create to report on and publicize project activities. In addition, as part of our outreach activities, we plan to conduct a number of workshops, short courses, tutorials, and similar activities, and the content of all of these will be available on the website. We intend to make any relevant course materials (and, hopefully, recorded lectures) available through our website.